



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Substitutions in the glycogenin-1 gene are associated with the evolution of endothermy in sharks and tunas**

**Citation for published version:**

Ciezarek, AG, Dunning, LT, Jones, CS, Noble, LR, Humble, E, Stefanni, SS & Savolainen, V 2016, 'Substitutions in the glycogenin-1 gene are associated with the evolution of endothermy in sharks and tunas' *Genome Biology and Evolution*, vol. 8, no. 9, pp. 3011-3021. DOI: 10.1093/gbe/evw211

**Digital Object Identifier (DOI):**

[10.1093/gbe/evw211](https://doi.org/10.1093/gbe/evw211)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

*Genome Biology and Evolution*

**Publisher Rights Statement:**

The Author(s) 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Substitutions in the Glycogenin-1 Gene Are Associated with the Evolution of Endothermy in Sharks and Tunas

Adam G. Ciezarek<sup>1</sup>, Luke T. Dunning<sup>1,2</sup>, Catherine S. Jones<sup>3</sup>, Leslie R. Noble<sup>3</sup>, Emily Humble<sup>1,4</sup>, Sergio S. Stefanni<sup>5</sup>, and Vincent Savolainen<sup>1,\*</sup>

<sup>1</sup>Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, UK

<sup>2</sup>Present address: Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

<sup>3</sup>Institute of Biological and Environmental Sciences, School of Biological Sciences, University of Aberdeen, Zoology Building, Tillydrone Avenue, Aberdeen, Scotland, UK

<sup>4</sup>Present address: Department of Animal Behaviour, University of Bielefeld, Postfach 100131, Bielefeld, Germany

<sup>5</sup>Villa Comunale, Stazione Zoologica Anton Dohrn, Naples, Italy

\*Corresponding author: E-mail: v.savolainen@imperial.ac.uk.

Accepted: August 25, 2016

**Data deposition:** BioProject number: PRJNA305977; SAMN04992863: Mako shark; SAMN04992864: Porbeagle; SAMN04992865: Basking shark; SAMN04992866: Atlantic bluefin tuna; SAMN04992867: Southern bluefin tuna; SAMN04992868: Bigeye tuna; SAMN04992869: Yellowfin tuna; SAMN04992870: Albacore tuna; SAMN04992871: Skipjack tuna; SAMN04992872: Atlantic mackerel.

## Abstract

Despite 400–450 million years of independent evolution, a strong phenotypic convergence has occurred between two groups of fish: tunas and lamnid sharks. This convergence is characterized by centralization of red muscle, a distinctive swimming style (stiffened body powered through tail movements) and elevated body temperature (endothermy). Furthermore, both groups demonstrate elevated white muscle metabolic capacities. All these traits are unusual in fish and more likely evolved to support their fast-swimming, pelagic, predatory behavior. Here, we tested the hypothesis that their convergent evolution was driven by selection on a set of metabolic genes. We sequenced white muscle transcriptomes of six tuna, one mackerel, and three shark species, and supplemented this data set with previously published RNA-seq data. Using 26 species in total (including 7,032 tuna genes plus 1,719 shark genes), we constructed phylogenetic trees and carried out maximum-likelihood analyses of gene selection. We inferred several genes relating to metabolism to be under selection. We also found that the same one gene, glycogenin-1, evolved under positive selection independently in tunas and lamnid sharks, providing evidence of convergent selective pressures at gene level possibly underlying shared physiology.

**Key words:** tuna, sharks, endothermy, positive selection, phylogenetics.

## Introduction

Bony fishes and sharks have been separated by up to 450 million years of independent evolution (Venkatesh et al. 2014). As such they differ in many aspects of their physiology, anatomy, and biochemistry (Bernal et al. 2001). Despite this, there is remarkable phenotypic convergence between two groups of active, epipelagic predators: the lamnid sharks (=family Lamnidae) and the tunas (=genera *Thunnus*, *Euthynnus*, *Auxis*, and *Katsuwonus* within family Scombridae). These two groups have a distinctive positioning

of red muscle (RM), a specialized swimming style and can warm up regions of body (endothermy).

First, RM is used for slow-twitch contraction, as in steady state swimming. It is primarily fuelled by aerobic metabolism, and as such is relatively rich in mitochondria and myoglobin compared to white muscle (WM). This WM is fast-twitch muscle used for burst swimming, primarily fuelled by glycolysis (anaerobically). As such, WM has lower concentrations of myoglobin and mitochondria (Dickson 1996). Typically, in fish, the majority of RM is located superficially, close to the outside of

the body. In contrast, in tunas and lamnid sharks, the majority of RM is located in a more central position within the body (Block and Finnerty 1994; Bernal et al. 2001). The centralization of RM has been directly associated with “thunniform” swimming and regional endothermy.

Second, “thunniform” swimming is characterized by the restriction of lateral movements to the caudal region (Donley et al. 2004; Gemballa et al. 2006). Force generated by the RM is transmitted efficiently to the tail, without causing local bending of a stiffened body (Westneat et al. 1993; Syme and Shadwick 2011). The RM is also a major source of metabolically generated heat. The evolution of regional endothermy requires a mechanism of insulation. In both groups, this is achieved using counter-current heat exchangers, which surround the centralized RM, enabling the maintenance of an elevated body temperature (Block and Finnerty 1994).

Third, endothermy is generally associated with high metabolic rates and high aerobic capacities (Nespolo et al. 2011). However, measuring metabolic rates in large, active fish is challenging (Blank et al. 2007). It is frequently stated that tunas have high mass-specific standard and maximum metabolic rates compared to ectothermic fish (Dickson and Graham 2004; Qiu et al. 2014). Measures of shark metabolic rates are mostly lacking, although high values have been recorded in the endothermic mako shark, *Isurus oxyrinchus* (Bernal et al. 2012). Measurements of the activity of metabolic enzymes have indicated that tuna’s WM have not only elevated anaerobic capacity, but also aerobic capacity, compared to ectothermic Scombridae (Dickson 1996). The same pattern has also been demonstrated in lamnid sharks (Bernal, Smith, et al. 2003). The elevated aerobic capacity of WM may enable rapid repayment of the oxygen debt induced by burst swimming, increasing speed of recovery (Korsmeyer and Dewar 2001; Bernal and Sepulveda 2005).

The genetic mechanisms underlying phenotypic convergence between laminid sharks and tunas are unknown. Although metabolic pathways are highly conserved across eukaryotes (Fernie et al. 2004), positive selection has been detected in enzymes of taxa under strong selective pressure for metabolic performance, such as consumption of very large prey in snakes or cold adaptation in insects (Castoe et al. 2008; Dunning et al. 2013). Lamnid sharks and tunas are also under strong pressure. Thunniform swimming and regional endothermy are associated with a high metabolic cost (Watanabe et al. 2015). This is particularly problematic due to the nutrient-poor pelagic environment these fish occupy (Korsmeyer et al. 1996) and the high rate of thermal diffusion in water, making heat retention difficult (Carey et al. 1971). These traits enable an increased thermal range (Dickson and Graham 2004) and cruise swimming speed (Watanabe et al. 2015) in RM endotherms and could therefore provide a strong selective benefit under conditions where this enables better access to high-energy prey (Madigan et al. 2015).

Phylogenetic approaches for detecting evidence of positive selection can be used to nominate candidate genes. Therefore here, we test two hypotheses: (i) given the elevated metabolic capacities of WM in endotherms, genes associated with muscle metabolism in lamnid sharks and tunas would have evolved under positive selection; and (ii) given convergent evolution between tunas and lamnid sharks, we expect to find orthologous genes involved in muscle metabolism to be under positive selection in both groups. To test these hypotheses, we sampled the WM of a range of endothermic tunas and sharks along with their closest ectothermic relatives. We sequenced their WM transcriptomes, which we also supplemented with published RNA-seq data. We then applied comparative phylogenetic analyses search for candidate genes for selection, which may underlie the phenotypic convergences observed in lamnid sharks and tunas.

## Materials and Methods

### Sampling

WM samples of two lamnid sharks and one ectothermic species (basking shark; table 1), and seven Scombridae species, including six endothermic tunas, and ectothermic mackerel were collected between January 2013 and February 2014. This data was supplemented by previously published RNA-seq data for 11 species (table 1). All samples were stored at  $-20^{\circ}\text{C}$  in RNeasy Lysis Buffer (Qiagen, Crawley, UK). Prior to RNA extraction, all samples were disrupted and homogenized using the Powergen homogeniser (Fisher Scientific, Loughborough, UK). Total RNA was extracted and purified using the RNeasy Fibrous Tissue Mini Kit and MiniElute Cleanup Kit (Qiagen, Venlo, Netherlands) following the manufacturer’s protocol. RNA quality and quantity were assessed using a Nanodrop ND2000 (Nanodrop Technologies, Wilmington, DE), a TAE-agarose gel and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

In order to verify species identity of fish purchased from traders, we sequenced cytochrome *b* (*cyt b*) (Botti and Giuffra 2010). A cDNA reverse transcription kit (Applied Biosciences Inc, Foster City, CA) was used to generate cDNA following the manufacturers protocol. PCR amplifications were carried out using a RedTaq ReadyMix PCR Reaction Mix (Sigma-Aldrich, St. Louis MO) using primers adapted from Botti and Giuffra (2010) (supplementary table S1, Supplementary Material online) and a Veriti Thermal Cycler (Applied Biosciences Ltd, Foster City CA). PCR products were then purified with ExoSAP-IT (Affymetrix Inc, Santa Clara, CA) and sequenced using Big Dye Terminator v3 (Applied Biosciences Inc, Foster City, CA). Sequencing product was subsequently cleaned using ethanol and sodium acetate precipitation, and run on a 3130xl Genetic Analyzer (Applied Biosciences Inc, Foster City CA). Electropherograms were edited using Geneious (v6) and blastn-searched against GenBank.

**Table 1**The origin of samples used for this study and *de novo* trinity assembly statistics

Common name	Species name	Origin	Paired-end reads used for assembly (million)	Number of assembled contigs	Contig N50	Number of coding regions after clustering
Yellowfin tuna	<i>Thunnus albacares</i>	Purchased, UK	57.9	61,045	1,851	18,343
Atlantic bluefin tuna	<i>Thunnus thynnus</i>	Purchased, UK	58	76,764	1,593	21,922
Bigeye tuna	<i>Thunnus obesus</i>	Purchased, UK	59.9	74,882	1,851	21,066
Skipjack tuna	<i>Katsuwonus pelamis</i>	Azores	59.3	83,724	1,414	20,385
Southern bluefin tuna	<i>Thunnus maccoyii</i>	Australia	53.9	58,944	1,017	15,170
Albacore tuna	<i>Thunnus alalunga</i>	Australia	53.8	81,372	1,967	22,263
Atlantic mackerel	<i>Scomber scombrus</i>	Purchased, UK	58.2	65,763	761	15,335
Atlantic bonito	<i>Sarda sarda</i>	Assembly kindly provided by the authors of Sarropoulou et al. (2014)	162.1	68,220	3,011	27,010
Pacific bluefin tuna	<i>Thunnus orientalis</i>	Supplementary information of Yasuike et al. (2016)	–	40,813	1,722	28,471
Black scabbardfish	<i>Aphanopus carbo</i>	Stefanni et al. (2014)	–	8,319	619	1,055
Yellowtail kingfish	<i>Seriola lalandi</i>	SRR2138320	95.7	138,558	2,204	34,218
Barramundi	<i>Lates calcarifer</i>	GAQL01000001.1-01363785.1	–	363,785	1,680	54,776
Porbeagle	<i>Lamna nasus</i>	UK	16.6	53,103	708	8,694
Shortfin mako shark	<i>Isurus oxyrinchus</i>	Azores	81.4	81,680	892	15,046
Great White shark	<i>Carcharodon carcharias</i>	ORFs taken from Richards et al. (2013)	–	105,313	640	17,134
Sandtiger shark	<i>Carcharias taurus</i>	SAMN03333352	71.7	118,363	1,687	24,769
Basking shark	<i>Cetorhinus maximus</i>	UK	61.5	19,017	343	1,630
Smooth dogfish	<i>Mustelus canis</i>	SAMN03333350	52.7	98,463	2,026	19,990
Lemon shark	<i>Negaprion brevirostris</i>	SAMN03333351	62.3	70,506	1,701	16,217
Caribbean reef shark	<i>Carcharhinus perezi</i>	SAMN03333349	62	111,848	2,340	23,075
Bull shark	<i>Carcharhinus leucas</i>	SAMN03333348	60.5	91,122	1,719	21,657
Blue shark	<i>Prionace glauca</i>	SAMN03333347	65.8	96,740	1,137	17,669
Tiger shark	<i>Galeocerdo cuvier</i>	SAMN03333353	59.1	179,867	1,858	26,843
Atlantic sharpnose shark	<i>Rhizoprionodon terraenovae</i>	SAMN03333345	60.5	88,870	1,844	19,646
Small-spotted catshark	<i>Scyliorhinus canicula</i>	http://skatebase.org; last accessed March 2014	–	107,231	695	24,218
Blacknose shark	<i>Carcharhinus acronotus</i>	SAMN03333346	57.8	131,575	2,201	22,956

### Construction and Sequencing of cDNA Libraries

We commissioned 3'-fragment normalized cDNA libraries for construction by an external company (BGI Tech Solutions, Hong Kong). Using The TruSeq RNA Library Preparation Kit v2, cDNA libraries were produced with DSN normalization. These normalized cDNA libraries were then sequenced using Illumina HiSeq 2000 (Illumina Inc, San Diego, CA). Initial quality control was carried out by BGI Tech Solutions, with low quality reads (phred score <20) removed and primer and adaptor sequences trimmed. Upon retrieval, cleaned reads were evaluated using FastQC (v0.10.1), and then assembled into contigs using Trinity (v2013-08-05) (Grabherr et al. 2011), with default settings. Data are available on GenBank under Bioproject number PRJNA305977.

### Gene Prediction and Annotation

For each transcript, the longest open reading frame (ORF) was extracted using TransDecoder [trinity package: (Grabherr et al.

2011)]. Stop codons as well as contigs that returned more than one TransDecoder ORF were removed from the dataset. To reduce redundancy, each set of ORFs was clustered using CD-HIT-EST using a cut-off of 0.98 (Fu et al. 2012). A phylogenetic tree-based approach was then used to detect orthologs between the sharks and between the perciform fish separately. Clustered ORF assemblies were translated. To guide orthogroup assignment, cDNA sequences for *Danio rerio*, *Homo sapiens*, *Mus musculus*, *Latimeria chulmnae*, and *Oryzias latipes* were downloaded from the Ensembl database (Yates et al. 2015). Using TransDecoder, coding sequences were extracted from each contig. These contigs were then clustered using CD-HIT-EST using a cut-off of 0.98, and translated. Orthofinder was used to infer homolog groups (Emms and Kelly 2015). The first step of this was an all-versus-all blastp (v2.2.25) search (Altschul et al. 1990). Orthofinder then normalizes blast scores for sequence length and phylogenetic distance before selecting putative gene pairs for orthogroup inference using the MCL clustering algorithm



(Van Dongen 2000). Following the methods and scripts of Yang and Smith (2014), we then trimmed orthogroup trees. First, terminal branches with an absolute length of 2, or relative length of 10 times that of their sister were trimmed. As RNA-seq data includes multiple splice variants and isoforms, monophyletic or paraphyletic groups can arise from the same species. In these instances, the contig with the highest number of aligned characters was retained, with the remainders trimmed. Deep paralogs, with a branch length of greater than 0.5, were then cut. A raw coding sequence file was then generated for each orthogroup. This was re-aligned using mafft (v7.2.45) (Kato and Standley 2013), and phylogenetic trees inferred using RAXML (v8.1.17) (Stamatakis 2006). Following a repeat of the trimming procedure, orthologs were inferred using the “prune\_paralogs\_RT.py” script of Y. Yang and Smith (2014). This RT method explicitly accounts for gene duplications. *Danio rerio*, *H. sapiens*, *M. musculus*, *L. chulmnae*, and *O. latipes* were used here as outgroups to root the trees, but then were trimmed. For downstream analyses, we only looked at putative orthologs that were identified across at least five species and included at least two endothermic and two ectothermic species.

Alignment error has been demonstrated to be a key source of false positives in positive selection inferences (Markova-Raina and Petrov 2011; Redelings 2014). To reduce the likelihood of this, we used a stringent alignment approach. The putatively orthologous nucleotide sequences were first translated to proteins. Using m-coffee, implemented within the tcoffee v11 software package (Notredame et al. 2000), these amino acid sequences were aligned using four separate aligners: muscle\_msa, mafft\_gins\_msa, tcoffee\_msa and kalign\_msa. Output scores were given for each alignment site based on the concordance of the different aligners. All sites with a concordance less than nine, which indicates total concordance, were trimmed. Regions in the resultant alignments that are highly divergent may not be truly orthologous, or still may be influenced by alignment, sequencing or assembly error. To further control for this, alignment quality of each column was analyzed using the Transitive Consistency Score (TCS) alignment evaluation score implemented within t-coffee (Chang et al. 2014). Only columns with the maximum quality score of nine were retained. These trimmed protein sequences were back-translated to their corresponding nucleotides. Codons absent in at least half of the species were removed. Maxalign v1.1 (Gouveia-Oliveira et al. 2007) was then used to detect and remove poorly aligned gap-rich sequences. This reduces the risk of paralogous sequences being analyzed, as these are less likely to align well. All sequences which had sequences discarded by maxalign were then checked to see if the remaining good quality aligned sequences still considered enough species to be considered (using the same criteria outlined earlier). If this was the case, they were realigned and trimmed, with the low quality sequences removed, using the same method and put forward for analysis.

All trinity transcripts and ORFs corresponding to good quality alignments were annotated using the Trinity Trinotate pipeline (Grabherr et al. 2011). Sequences were searched against UniProt using the SwissProt and Uniref90 databases (E-value cutoff 1E–10). Coding sequences were also searched for conserved protein domains using Pfam (Finn et al. 2014). Additionally, ORFs were BLASTx-searched against NCBI nr database and annotated using Blast2go v2.5 (Conesa et al. 2005).

### Phylogenetic Inferences

Analyses for positive selection using the PAML software require an accurate phylogenetic tree (Yang 2007). Separate phylogenetic trees were inferred for the sharks and bony fish (including tunas). In each case, 4-fold degenerate sites were extracted from each putative ortholog, and concatenated to produce a 4-fold supermatrix. This ensures that phylogeny reconstruction was independent from detecting positive selection because 4-fold degenerate sites do not affect the sequence of amino acids in the translated protein. Using phyutility (v2.2.26), alignment columns with less than half of species present were trimmed. A maximum-likelihood phylogenetic tree was built for each dataset, using RAXML (v8.1.18), with 1,000 rapid bootstraps (Stamatakis 2006; Stamatakis et al. 2008). In each case, the model of evolution was determined using the best Akaike information criteria (AIC) scores (Posada and Buckley 2004), using jModeltest v2.1.10 (Posada 2008). For each dataset, a Bayesian phylogenetic tree was also inferred using ExaBayes (Aberer et al. 2014). Four independent MCMC runs, each with three coupled chains, were run for 1,000,000 generations, sampling every 500. Using the “sdsf” and “postProcParam” tools of exabayes, along with Tracer (v1.6) (Rambaut et al. 2013), we ensured that average deviation of split frequencies was close to zero, potential scale reduction factors were close to one and effective sample sizes of estimated parameters were greater than 200.

### Detecting Positive Selection

The CodeML programme of the PAML (v4.7) package was used to analyze all alignments for positive selection (Yang and Bielawski 2000). The branch-site test was implemented for each ortholog (Zhang et al. 2005). These models require the specification of a “foreground” branch, which can be tested for evidence of selection. As our hypothesis relates to selection underlying the evolution of endothermy, the root branch of the endothermic taxa was selected and subjected to a branch-site test implementing two models. One model allowed for selection with four site classes:  $0 < \omega < 1$  in both branch-classes,  $\omega = 1$  in both,  $\omega_{\text{foreground}} > 1 = \omega_{\text{background}}$  and  $\omega_{\text{foreground}} > 1 > \omega_{\text{background}}$ , in which  $\omega$  denotes the ratio of the number of non-synonymous substitutions per non-synonymous site ( $d_N$ ) to the number of synonymous

substitutions per synonymous site ( $d_s$ ). The null model differs in that  $\omega$  foreground cannot exceed 1. A likelihood ratio test (LRT) with  $\chi^2$  was then used to compare models, and test whether the model allowing for selection fits the data significantly better than the null model. As different genes contained different numbers of species, a newick-formatted tree file was generated for each individual gene. When taxa were absent, they were removed from a base tree with all taxa present, using Newick Utilities (Junier and Zdobnov 2010). Three runs with different starting  $\omega$  values (0.5, 1, 1.5) for each gene were carried out for each alignment.

A key source of false-positive in the branch-site model is alignment error (Fletcher and Yang 2010; Markova-Raina and Petrov 2011), which even with strict alignment procedures cannot be conclusively eliminated. All genes with LRT  $P < 0.05$  were considered putatively under selection and independently re-analyzed using a different alignment procedure. These orthologs were translated and aligned using PRANK v100802 (Löytynoja and Goldman 2005) implemented through Guidance v1.5 (Penn et al. 2010). PRANK has been demonstrated to produce a low rate of false-positives in branch-site tests compared to other alignment softwares (Markova-Raina and Petrov 2011; Redelings 2014). Guidance evaluates the quality of alignments using two methods: the heads or tails method, which evaluates uncertainty generated by co-optimal solutions and a Guidance method, which analyses alignment robustness to guide-tree uncertainty. These raw alignments were then back-translated to nucleotides. Protein alignment residues with a low score (using the default recommendation of  $\leq 0.93$ ) were removed. Alignments were then parsed to Trimal v1.3 (Capella-Gutiérrez et al. 2009) where columns with gaps in at least 40% of sequences or with a similarity score of  $< 0.001$  were trimmed. These trimmed protein alignments were then used to trim the nucleotide alignment. As with the original alignments, these were assessed using Maxalign and realigned without excluded taxa if necessary. These alignments were then analyzed with the branch-site test implemented through a different software, slimcodeml (v2013 – 02 – 07). As with the t-coffee alignments, each gene was analyzed with the same three different starting  $\omega$  values. For each ortholog, all  $P$  values generated by separate branch-site tests (i.e., between the different alignment methods and starting  $\omega$ ) were corrected using the method of Benjamini–Hochberg (Benjamini and Hochberg 1995). We only considered genes to be under selection if  $P$  values were  $< 0.05$  (Benjamini–Hochberg adjusted for all tests on the same gene) for all six tests carried out. Genes found to be under-selection were tested for enrichment, using all the genes analyzed as the background. A Fisher's exact test, with Benjamini–Hochberg adjustment for multiple comparisons, was implemented in Blast2go v2.8.0 (Conesa et al. 2005), based on the gene ontology (GO) annotations from the nr database. They were also analyzed using the CodeML free-ratio and one-ratio tests. Genes with an

overall  $d_s > 1$  in the one-ratio test, or  $d_s > 1$  in the endothermic root in the free-ratio test, were inferred to be influenced by synonymous-site saturation. This may influence the reliability of the branch-site test (Gharib and Robinson-Rechavi 2013; Roux et al. 2014).

Genes with the same gene annotation found to be under selection in both the lamnid sharks and tunas were investigated further. To assess whether these genes have a particularly overall high rate of  $d_N$  or  $d_N/d_S$ , free-ratio and one-ratio tests were carried out in CodeML. High rates of non-synonymous mutations may make a gene a more likely target of positive selection, as there is an increased chance of an advantageous allele arising (Montoya-Burgos 2011). To assess whether orthology inference was accurate, coding sequences from genome projects on the corresponding gene tree were downloaded from Ensembl (Yates et al. 2015). These were aligned with the corresponding orthologs from our dataset using mafft (v7.245). The gene tree was then constructed using RAxML with 200 rapid bootstraps (v8.1.17).

### Ancestral State Reconstructions

For genes inferred to be under selection in both tunas and lamnid sharks, we used ancestral state reconstructions, inferring specific amino acid substitutions using FastML (v3.1) (Ashkenazy et al. 2012). We then compared the ancestral amino acid sequence of the endotherms to the ancestral sequences of their closest ectotherms. We visualized these changes on a human high-resolution crystal structures (Chaikuad et al. 2011), downloaded from the Protein Data Bank in Europe (Velankar et al. 2010).

## Results and Discussion

### Orthologous Genes

Between 16.6 and 81.4 million paired-end reads were sequenced for each species (mean 58 million; table 1). Using tree-based orthology detection techniques, an initial 12,982 bony fish and 6,620 shark orthologs were detected. Following the t-coffee alignment pipeline, 7,798 and 2,086 bony fish and shark orthologs remained for phylogenetic analyses. To detect positive selection using PAML, we needed orthologs present in the sisters to tunas and lamnid sharks (namely, Atlantic bonito and sandtiger shark), which comes down to 7,032 bony fish and 1,719 shark genes (supplementary table S2, Supplementary Material online).

### Phylogenetic Trees

We built two supermatrices consisting of all of the 4-fold degenerate sites from the 7,798 and 2,086 orthologs above. This resulted in a matrix of 701,592 nucleotides for bony fish and 173,967 for sharks. The "GTRGAMMA" model of evolution was used in each maximum-likelihood inference. The percentage of gaps for each species ranged from 10.1% to 97% in

bony fish and 8.2% to 89.7% in sharks (supplementary table S3, Supplementary Material online).

We recovered a strongly supported phylogenetic tree for sharks, with all branches receiving 99–100% bootstrap and posterior probability of 1.0 (fig. 1a). This agreed with recent studies that used five genes (Vélez-Zuazo and Agnarsson 2011), but not with Sorenson et al. (2014) in placing *Carcharias taurus* as the closest relative to the endothermic sharks (rather than *Cetorhinus maximus*). Notably, data of *C. maximus* were available only in 10.3% of our matrix. The sample came from an individual stranded dead on a beach, which may have reduced RNA quality. Only 1,630 clustered, filtered coding sequences were analyzed for this species, compared to a mean of 18,593 for the other sharks. However, all nodes received bootstrap support 100% and posterior probability of 1.0, indicating that there was still sufficient data for *C. maximus*.

The bony fish dataset also produced a strongly supported tree (fig. 1b). Although *Aphanopus carbo* was poorly represented in the matrix (data present only for 3.0% of the aligned sites), it was confidently placed in a position consistent with previously published trees (Miya et al. 2013). Relationships amongst the *Thunnus* species were in agreement with a recently published RAD-seq based phylogenetic tree (Díaz-Arce et al. 2016).

### Positive Selection

After the two alignment procedures (see “Methods” section), 139 genes (1.9%) were inferred to be under positive selection in tunas, and 19 (1.1%) in the lamnid sharks (supplementary table S4, Supplementary Material online). No evidence for GO term enrichment was found in either dataset. We found evidence of synonymous-site saturation in three genes inferred to be under selection in the lamnid sharks (saturation inferred from an overall  $d_s > 1$ : *MYG*: 1.14, *BTNL1*: 1.34 and an unidentified protein: 3.7897). However, there was no significant evidence of enrichment of genes inferred to be under selection with overall  $d_s > 1$  compared to all genes tested (Fisher’s exact test,  $P = 0.28$ ). We found evidence of saturation in ten of the tuna genes (overall  $d_s > 1$ : *AATC*: 1.08, *TNF6CB*: 1.50, *RIR1*: 1.00, *LYG*: 1.00, *IRF8*: 1.00, *CD37*: 1.21, *COPT1*: 1.20, *PBDC1*: 1.09, *RN214*: 1.34, *MYSM1*: 1.04). As with sharks, there was no significant enrichment (Fisher’s exact test,  $P = 0.64$ ) indicating that it was not a significant cause of false positives. Indeed, simulation studies have suggested that a high  $d_s$  causes a lack of power in the branch-site test rather than an excess of false-positives (Gharib and Robinson-Rechavi 2013).

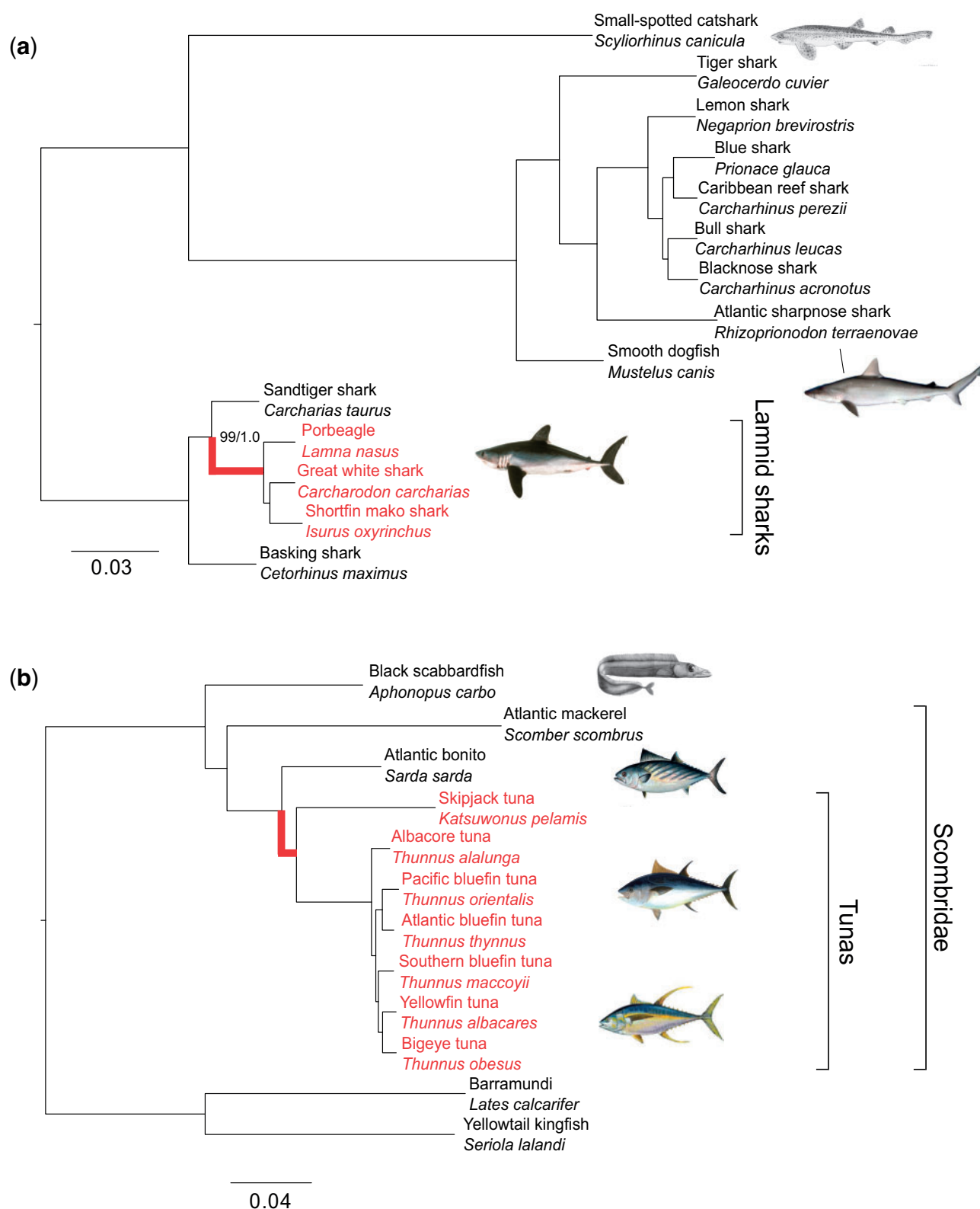
One gene was inferred to be under selection in both tunas and lamnid sharks independently: the glycogenin-1 gene (*GLYG1*) (supplementary table S4, Supplementary Material online). This is unlikely due to chance (we estimated the probability of a given gene been under selection in both groups

independently to be 0.0002, given that 1,192 genes with the same gene name were present in both groups, of which 14 were under selection in sharks and 22 in tunas). This crude estimation assumes that genes are equally likely to be under selection. To examine whether *GLYG1* is a particularly fast evolving gene, we performed free-ratio and one-ratio tests using CodeML, to test whether  $\omega$  is high either overall or in background branches. In bony fish, overall  $\omega$  for the *GLYG1* gene was 0.21, which was within one standard deviation (0.16) of the mean  $\omega$  of all bony fish genes tested (0.14). The free-ratio model revealed an  $\omega$  of 0.40 in the root of tunas and 0.83 at the root of lamnid sharks. All other branches with high values of  $\omega$  (branches with 1.06, 1.18, and 0.79 within the tunas, one with 999 within the sharks) were supported by low values of  $d_s$  ( $< 0.00001$ ), rather than elevated  $d_N$ , indicating low genetic differentiation rather than a fast-evolving gene. The overall  $\omega$  was 0.13 in the shark *GLYG1* gene. This was also within one standard deviation (0.16) of the mean  $\omega$  of all shark genes tested (0.17). These results indicate that *GLYG1* is not a particularly fast evolving gene in these taxa. However, sampling of this gene with greater phylogenetic coverage may provide further information as to whether it has undergone selection in separate groups.

Note that we did not have *GLYG1* data in the following taxa: *Katsuwonus pelamis*, *Thunnus alalunga*, *Thunnus obesus*, *A. carbo*, *Lates calcarifer*, *Seriola lalandi*, *Lamnias nasus*, *C. maximus*, and *Prionace glauca*.

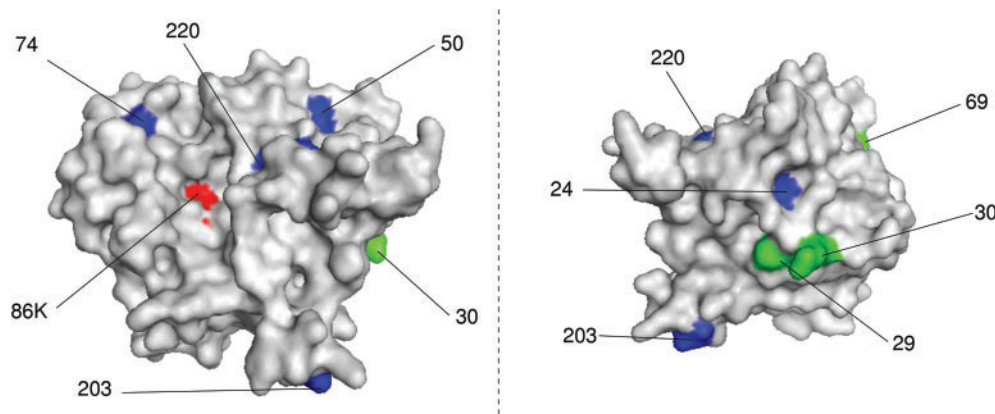
To confirm that we were analyzing true orthologs of *GLYG1*, we inferred a gene tree using *GLYG1* genes from our data sets, along with *GLYG1*, paralogs *GLYG1a*, *GLYG1b*, and *GLYG2* genes from other genomes. All our bony fish genes clustered within the *GLYG1a* teleost genes (supplementary fig. S1, Supplementary Material online). Similarly, all our shark genes clustered together as sister to coelacanth and tetrapod *GLYG1* genes (supplementary fig. S1, Supplementary Material online). These results support our ortholog inference rather than comparing multiple isoforms. Although there has been a gene duplication in *GLYG1* in bony fish (Yates et al. 2015), the isoform we analyzed here is *GLYG1a*.

*GLYG1* is a candidate gene for recovery following the predatory behavior of tunas and lamnid sharks. The glycogenin encoded by *GLYG1* is an enzyme involved with the muscular genesis of glycogen, which is particularly important in fast-twitch muscle (Cussó et al. 2003). The rate at which muscular glycogen is restored following exercise dictates how quickly an individual can recover from exercise. It has been demonstrated that the tuna *K. pelamis* can do this rapidly, at a rate similar to mammals (Weber et al. 1986). In humans, increased expression of *GLYG1* has been found during recovery from exercise (Kraniou et al. 2000), and has been associated with increased muscular glycogen content (Zhang et al. 2013). Mutations in this gene also result in glycogen depletion of the skeletal muscle (Nilsson et al. 2012). Although measurements of



**FIG. 1.**—Phylogenetic tree of sharks (a) and bony fish (b). Endothermic species as well as their root branch are in red. All nodes were fully supported (1.0 posterior probability, 100% bootstrap support) unless otherwise indicated. Scale bar refers to branch length (number of expected substitutions per site). Images taken from <http://en.wikipedia.org>, <http://commons.wikipedia.org>.





**Fig. 2.**—Two views of the structural modelling of human glycogenin-1, showing amino acids changes between endothermic and ectothermic fish (see text for details): changes endothermic sharks in blue, changes in endothermic tunas in green, active site in red Lysine 86; other numbers refer to amino acid position in the human protein.

muscular glycogenin genesis have not been made in lamnid sharks, high activities of both lactate dehydrogenase (LDH) and citrate synthase (CS) have been documented compared to their ectothermic relatives (Bernal, Smith, et al. 2003). These are markers of anaerobic and aerobic metabolic capacity, respectively (Dickson 1996). The relative abundance of these enzymes has also been found to positively correlate, and so it has been speculated that endothermic sharks are able to clear lactate in a similar manner to tunas (Bernal and Sepulveda 2005). This suggests a similarly elevated rate of exercise recovery.

We also inferred the ancestral sequences of *GLYG1* in our trees. Mapping amino acid changes on human *GLYG1* shows that these are at the surface of the protein, with the exception of one change in tunas (fig. 2). Such surface sites are likely to influence thermal stability of the enzyme (Fields 2001), although not excluding an effect on catalytic performance (Fields and Somero 1998).

Given the large genome size of sharks (Venkatesh et al. 2014) and evolutionary distance between sharks and tunas, it is not surprising that we did not find substitutions at convergent amino acid sites within *GLYG1*. Convergent evolution is expected to be more common in organisms with small genomes, as there are fewer mutational target sites which could influence fitness (Stern 2013). Additionally, the fitness effects of substitutions are dependent on the genetic background. The great evolutionary distance between tunas and sharks is likely to have reduced the probability of parallel or convergent substitutions at the same sites (Storz 2016).

We also found evidence of selection in other genes associated with metabolism (supplementary table S4, Supplementary Material online). This included one electron transport chain gene (*COX41*) in lamnid sharks. This gene was tested, but not inferred to be under selection in the tunas. Four lipid metabolism genes (*MCAT*, *ACOT1*, *ACOT4*,

and *ACOT13*) were inferred to be under selection in tunas along with two genes associated with glycolysis (*TPISB*, *TIGRA*). Of these genes, only *ACOT4* was tested in sharks, and was not inferred to be under selection.

The electron transport chain gene inferred to be under selection in the sharks, *COX41*, encodes a subunit of cytochrome-c. This is the last enzyme in the electron transport chain, and plays a key role in aerobic respiration (Wikström 2010). In tunas, three *ACOT* (Acyl-CoA-synthetases) were inferred to be under selection. Acyl-CoA-synthetases facilitate  $\beta$ -oxidation, by providing CoA (Hunt et al. 1999). *MCAT* also may play a role in facilitating  $\beta$ -oxidation. Mitochondrial carnitine/acylcarnitine carrier proteins catalyse transport of acyl-carnitine into the mitochondria, increasing fatty acyl units in the mitochondrial matrix, where  $\beta$ -oxidation enzymes oxidize them (Indiveri et al. 2011).  $\beta$ -oxidation is vital to overall production of metabolic energy, where fatty acids are broken down to form acetyl coenzyme A, which enters the tricarboxylic acid cycle and feeds aerobic respiration (Indiveri et al. 2011). The protein encoded by *GPDA* plays a variety of metabolic roles, linking lipid metabolism to gluconeogenesis (Harding et al. 1975) and contributing electrons to the mitochondrial electron transport chain (Bunoust et al. 2005). This indicates adaptive evolution relating to aerobic metabolism occurred in the evolution of tunas and lamnid sharks, although not in the same pathways.

The protein encoded by *TPISB* operates at a branch-point influencing glycolytic flux (Compagno et al. 1999). As branch-point enzymes exhibit control over rate of glycolysis, such enzymes are likely to be a target of selection (Eanes 2011). *TIGRA* encodes a probable fructose-2,6-bisphosphatase. There is evidence that this controls phosphofructokinase-1 (a key glycolysis regulatory enzyme) and therefore the rate of glycolysis (Hue and Rider 1987). This indicates a selective pressure influencing glycolytic capacity in tunas, but not in sharks. Although higher

than other sharks, the metabolic capacities of lamnid shark WM are still lower than those of the tunas (Bernal, Smith, et al. 2003). This indicates that such adaptive evolution did not occur in the lamnid sharks, and that different mechanisms underlie its elevated glycolytic potential.

A further six genes with functions relevant to the physiology and behavior of tunas were inferred to be under selection (i.e., *MPSF*, *MYOZ2*, *LMOD3*, *RYR1*, and *MOT4*). These have functions relating to muscular contraction, muscular development and transmembrane lactate transport (van der Ven and Fürst 1997; Franzini-Armstrong 1999; Yuen et al. 2014; The UniProt Consortium 2015). Orthologs for these genes were not tested in sharks, and genes with similar functions were not inferred to be under selection. However, a myoglobin gene, *MYG*, was. Higher levels of myoglobin have been documented in the red muscle of endothermic sharks than ectothermic sharks, which would enhance diffusion of oxygen from the blood to the muscle cells (Bernal, Sepulveda, et al. 2003). *MYG* was tested, but not inferred to be under selection in the tunas.

## Conclusion

We hypothesized that selection would have acted on genes involved in metabolic pathways, as well as genes relating to muscular contraction and development. We found several such genes, including *COX41*, *TPISB*, *TIGRA*, *MCAT*, *ACOT1*, *ACOT4*, *ACOT13*, *MPSF*, *MYOZ2*, *LMOD3*, *RYR1*, and *MOT4*. We also hypothesized that the same genes will be found to be involving under positive selection in both lamnid sharks and tunas. We found this to be the case only for one gene, *GLY1*, which may have had a role in enhancing exercise recovery in the WM in each group. Further studies are needed to investigate how these amino acid substitutions are affecting the function of the enzyme.

Of course, the evolution of endothermy is more complex that involving just one gene. For example, ontogenetic studies in these fish should reveal the mechanisms underlying the centralization of RM, which is key to the phenotypic convergence in lamnid sharks and tunas. Further work, focusing on sequence evolution as well as gene expression in other tissues, such as RM, cardiac tissue, brain, and liver is needed. Our study was restricted to WM, in which not all of the individual's genes were expressed, and therefore not sequenced here. We actually used in our final dataset only 1,719 and 7,032 genes in the sharks and tunas, respectively, representing a relatively small proportion of the estimated 20,000–25,000 genes that must be present in these fish (Braasch et al. 2016). Increasing phylogenetic coverage is also likely to provide valuable insights. For example, the thresher shark, *Alopias vulpinus* may have evolved RM centralization and endothermy independently to lamnid sharks, but does not share the “thunniform” swimming or enhanced white muscle metabolic capacity (Bernal, Smith, et al. 2003; Bernal and

Sepulveda 2005). Similarly, the slender tuna, *Allothunnus fallai*, has centralized RM and regional endothermy, but its phylogenetic position is still unclear (Sepulveda et al. 2008; Santini et al. 2013); it was not sampled here.

There are few examples of gene convergence underlying the same trait in distantly related taxa. One famous example includes echolocation in bats and cetaceans, whose evolution was first reported to involve convergent changes in nearly 200 genes (Parker et al. 2013). It was subsequently documented that this did not exceed the background level of amino acid convergence between echolocating and non-echolocating lineages, even in hearing genes (Zou and Zhang 2015; Thomas and Hahn 2015). This exemplifies the difficulty to infer the genetic basis of complex traits. We hope we have contributed to elucidating some of the remarkable convergence between sharks and tunas, but we await further studies that look into a broader taxonomic coverage and detailed functional assays.

## Supplementary Material

Supplementary tables S1–S4 and figure S1 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

We thank Sean Tracey and Jaime McAllister for supplying albacore and southern bluefin tuna samples, Eva Giacomello for collecting the skipjack tuna sample, Elena Sarropoulou for providing the Atlantic bonito assembly, Helen Hipperson for assistance in the lab, Barbara Block and Ziheng Yang for advice, the editors and reviewers for comments, and the Leverhulme Trust and BBSRC for funding.

## Literature Cited

- Aberer AJ, Kobert K, Stamatakis A. 2014. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Mol. Biol. Evol.* 31:2553–2556.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Ashkenazy H, et al. 2012. FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* 40:580–584.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57:289–300.
- Bernal D, Smith D, et al. 2003. Comparative studies of high performance swimming in sharks II. Metabolic biochemistry of locomotor and myocardial muscle in endothermic and ectothermic sharks. *J. Exp. Biol.* 206:2845–2857.
- Bernal D, Carlson JK, Goldman KJ, Lowe CG, eds. 2012. *Biology of sharks and their relatives*. Boca Raton, FL: CRC Press.
- Bernal D, Dickson KA, Shadwick RE, Graham JB. 2001. Analysis of the evolutionary convergence for high performance swimming in lamnid sharks and tunas [Review]. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 129:695–726.
- Bernal D, Sepulveda C, Mathieu-Costello OA, Graham J. 2003. Comparative studies of high performance swimming in sharks I. Red

- muscle morphometrics, vascularization and ultrastructure. *J. Exp. Biol.* 206:2831–2843.
- Bernal D, Sepulveda CA. 2005. Evidence for temperature elevation in the aerobic swimming musculature of the common thresher shark *Alopias vulpinus*. *Copeia* 2005:163–168.
- Blank JM, Farwell CJ, Morrisette JM, Schallert RJ, Block BA. 2007. Influence of swimming speed on metabolic rates of juvenile pacific bluefin tuna and yellowfin tuna. *Physiol. Biochem. Zool.* 80:167–177.
- Block BA, Finnerty JR. 1994. Endothermy in fishes: a phylogenetic analysis of constraints, predispositions, and selection pressures. *Environ. Biol. Fishes* 40:283–302.
- Botti S, Giuffra E. 2010. Oligonucleotide indexing of DNA barcodes: identification of tuna and other scombrid species in food products. *BMC Biotechnol.* 10:1.
- Braasch I, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* 48:427–437.
- Bunoust O, Devin A, Avéret N, Camougrand N, Rigoulet M. 2005. Competition of electrons to enter the respiratory chain: a new regulatory mechanism of oxidative metabolism in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 280:3407–3413.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Carey FG, Teal JM, Kanwish JW, Lawson KD, Beckett JS. 1971. Warm-bodied Fish. *Am. Zool.* 11:137–143.
- Castoe TA, Jiang ZJ, Gu W, Wang ZO, Pollock DD. 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS One* 3:e2201.
- Chaikwad A, et al. 2011. Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proc. Natl Acad. Sci. U S A.* 108:21028–21033.
- Chang J-M, Di Tommaso P, Notredame C. 2014. TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Mol. Biol. Evol.* 31:1625–1637.
- Compagno C, Boschi F, Daleffe A, Porro D, Ranzi BM. 1999. Isolation, nucleotide sequence, and physiological relevance of the gene encoding triose phosphate isomerase from *Kluyveromyces lactis*. *Appl. Environ. Microbiol.* 65:4216–4219.
- Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.
- Cussó R, et al. 2003. Differences between glycogen biogenesis in fast- and slow-twitch rabbit muscle. *Biochim. Biophys. Acta.* 1620:65–71.
- Díaz-Arce N, Arrizabalaga H, Murua H, Irigoien X, Rodríguez-Ezpeleta N. 2016. RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. *Mol. Phylogenet. Evol.* 102:202–207.
- Dickson KA. 1996. Locomotor muscle of high performance fishes: What do comparisons of tunas with ectothermic sister taxa reveal? *Comp. Biochem. Physiol. A Physiol.* 113:39–49.
- Dickson KA, Graham JB. 2004. Evolution and consequences of endothermy in fishes. *Physiol. Biochem. Zool.* 77:998–1018.
- Donley JM, Sepulveda CA, Konstantidis P, Gemballa S, Shadwick RE. 2004. Convergent evolution in mechanical design of lamnid sharks and tunas. *Nature* 429:2617–2627.
- Dunning LT, et al. 2013. Positive selection in glycolysis among Australasian stick insects. *BMC Evol. Biol.* 13:215.
- Eanes WF. 2011. Molecular population genetics and selection in the glycolytic pathway. *J. Exp. Biol.* 214:165–171.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.
- Fernie AR, Carrari F, Sweetlove LJ. 2004. Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr. Opin. Plant Biol.* 7:254–261.
- Fields PA. 2001. Protein function at thermal extremes: balancing stability and flexibility [Review]. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 129:417–431.
- Fields PA, Somero GN. 1998. Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proc. Natl Acad. Sci. U S A.* 95:11476–11481.
- Finn RD, et al. 2014. Pfam: the protein families database. *Nucleic Acids Res.* 42:D222–D230.
- Fletcher W, Yang Z. 2010. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol. Biol. Evol.* 27:2257–2267.
- Franzini-Armstrong C. 1999. The sarcoplasmic reticulum and the control of muscle contraction. *FASEB J.* 13:S266–S270.
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152.
- Gemballa S, Konstantinidis P, Donley JM, Sepulveda C, Shadwick RE. 2006. Evolution of high-performance swimming in sharks: transformations of the musculotendinous system from subcarangiform to thunniform swimmers. *J. Morphol.* 267:477–493.
- Gharib WH, Robinson-Rechavi M. 2013. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Mol. Biol. Evol.* 30:1675–1686.
- Gouveia-Oliveira R, Sackett PW, Pedersen AG. 2007. MaxAlign: maximizing usable data in an alignment. *BMC Bioinformatics* 8:312.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–652.
- Harding W, Pyeritz A, Copeland S, White B. 1975. Role of glycerol 3-phosphate dehydrogenase in glyceride metabolism. Effect of diet on enzyme activities in chicken liver. *Biochem. J.* 146:223–229.
- Hue L, Rider MH. 1987. Role of fructose 2, 6-bisphosphate in the control of glycolysis in mammalian tissues. *Biochem. J.* 245:313.
- Hunt MC, et al. 1999. Peroxisome proliferator-induced long chain acyl-CoA thioesterases comprise a highly conserved novel multi-gene family involved in lipid metabolism. *J. Biol. Chem.* 274:34317–34326.
- Indiveri C, et al. 2011. The mitochondrial carnitine/acylcarnitine carrier: function, structure and physiopathology. *Mol. Aspects Med.* 32:223–233.
- Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* 26:1669–1670.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.
- Korsmeyer KE, Dewar H. 2001. Tuna Metabolism and Energetics. In: Tuna: Physiology Ecology and Evolution. In: Block BA, Stevens DE, editors. Fish physiology. Vol. 19. New York: Academic Press, p. 35–78.
- Korsmeyer KE, Dewar H, Lai NC, Graham JB. 1996. The aerobic capacity of tunas: adaptation for multiple metabolic demands. *Comp. Biochem. Physiol. Part A Physiol.* 113:17–24.
- Kraniou Y, Cameron-Smith D, Misso M, Collier G, Hargreaves M. 2000. Effects of exercise on GLUT-4 and glycogenin gene expression in human skeletal muscle. *J. Appl. Physiol.* 88:794–796.
- Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl Acad. Sci. U S A.* 102:10557–10562.
- Madigan DJ, et al. 2015. Assessing niche width of endothermic fish from genes to ecosystem. *Proc. Natl Acad. Sci. U S A.* 112:8350–8355.

- Markova-Raina P, Petrov D. 2011. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Res.* 21:863–874.
- Miya M, et al. 2013. Evolutionary origin of the Scombridae (tunas and mackerels): members of a paleogene adaptive radiation with 14 other pelagic fish families. *PLoS One* 8:e73535.
- Montoya-Burgos JL. 2011. Patterns of positive selection and neutral evolution in the protein-coding genes of Tetraodon and Takifugu. *PLoS One* 6:e24800.
- Nespolo RF, Bacigalupe LD, Figueroa CC, Koteja P, Opazo JC. 2011. Using new tools to solve an old problem: the evolution of endothermy in vertebrates. *Trends Ecol. Evol.* 26:414–423.
- Nilsson J, et al. 2012. Molecular pathogenesis of a new glycogenosis caused by a glycogenin-1 mutation. *Biochim. Biophys. Acta.* 1822:493–499.
- Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302:205–217.
- Parker J, et al. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* 502:228–231.
- Penn O, et al. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res.* 38:W23–W28.
- Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25:1253–1256.
- Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53:793–808.
- Qiu F, Kitchen A, Burleigh JG, Miyamoto MM. 2014. Scombroid fishes provide novel insights into the trait/rate associations of molecular evolution. *J. Mol. Evol.* 78:338–348.
- Rambaut A, Suchard M, Drummond AJ. 2013. Tracer v1.6. Available from: <http://tree.bio.ed.ac.uk/software/tracer/>; last accessed April 1, 2016.
- Redelings B. 2014. Erasing errors due to alignment ambiguity when estimating positive selection. *Mol. Biol. Evol.* 31:1979–1993.
- Richards VP, Suzuki H, Stanhope MJ, Shiji MS. 2013. Characterization of the heart transcriptome of the white shark (*Carcharodon carcharias*). *BMC Genomics* 14:697.
- Roux J, et al. 2014. Patterns of positive selection in seven ant genomes. *Mol. Biol. Evol.* 31:1661–1685.
- Santini FG, Carnevale G, Sorenson L. 2013. First molecular scombrid time-tree (Percormorpha: Scombridae) shows recent radiation of tunas following invasion of pelagic habitat. *Ital. J. Zool.* 80:210–221.
- Sarropoulou E et al 2014. The Atlantic bonito (*Sarda sarda*, Bloch 1973) transcriptome and detection of differential expression during larvae development. *PLoS One* 9:e87744.
- Sepulveda CA, Dickson KA, Bernal D, Graham JB. 2008. Elevated red myotomal muscle temperatures in the most basal tuna species, *Allothunnus fallai*. *J. Fish Biol.* 73:241–249.
- Sorenson L, Santini F, Alfaro ME. 2014. The effect of habitat on modern shark diversification. *J. Evol. Biol.* 27:1536–1548.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* 57:758–771.
- Steffani S et al 2014. Transcriptome of the Deep-Sea Black Scabbardfish, *Aphanopus carbo* (Perciformes: Trachuridae): Tissue-Specific Expression Patterns and characterization of depth-related functional genes. *Int. J. Genomics.* 2014:Article ID 267482.
- Stern DL. 2013. The genetic causes of convergent evolution. *Nat. Rev. Genet.* 14:751–764.
- Storz JF. 2016. Causes of molecular convergence and parallelism in protein evolution. *Nat. Rev. Genet.* 17:239–250.
- Syme DA, Shadwick RE. 2011. Red muscle function in stiff-bodied swimmers: there and almost back again. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366:1507–1515.
- The UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43:D204–D212.
- Thomas GWC, Hahn MW. 2015. Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. *Mol. Biol. Evol.* 32:1232–1236.
- Velankar S, et al. 2010. PDBE: Protein Data Bank in Europe. *Nucleic Acids Res.* 38:D308–D317.
- Vélez-Zuazo X, Agnarsson I. 2011. Shark tales: a molecular species-level phylogeny of sharks (Selachimorpha, Chondrichthyes). *Mol. Phylogenet. Evol.* 58:207–217.
- van der Ven PF, Fürst DO. 1997. Assembly of titin, myomesin and M-protein into the sarcomeric M band in differentiating human skeletal muscle cells in vitro. *Cell Struct. Funct.* 22:163–171.
- Van Dongen S. 2000. A Cluster algorithm for graphs. *Rep. Inf. Syst.* 10:1–40.
- Venkatesh B, et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505:174–179.
- Watanabe YY, Goldman KJ, Caselle JE, Chapman DD, Papastamatiou YP. 2015. Comparative analyses of animal-tracking data reveal ecological significance of endothermy in fishes. *Proc. Natl Acad. Sci. U S A.* 112:6104–6109.
- Weber JM, Brill RW, Hochachka PW. 1986. Mammalian metabolic flux rates in a teleost—lactate and glucose turnover in tuna. *Am. J. Physiol.* 250:R452–R458.
- Westneat MW, Hoese W, Pell CA, Wainwright SA. 1993. The horizontal septum: mechanisms of force transfer in locomotion of scombrid fishes (Scombridae, Perciformes). *J. Morphol.* 217:183–204.
- Wikström M. 2010. Cytochrome c Oxidase. *Encycl. Life Sci.* 1–10.
- Yang Y, Smith SA. 2014. Orthology inference in non-model organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol.* 31:3108–3092.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.
- Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15:496–503.
- Yasuie M, et al. 2016. A functional genomics tool for the Pacific bluefin tuna: development of a 44K oligonucleotide microarray from whole-genome sequencing data for global transcriptome analysis. *Gene* 576:603–609.
- Yates A, et al. 2015. Ensembl 2016. *Nucleic Acids Res.* 44:D710–D716.
- Yuen M, et al. 2014. Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. *J. Clin. Invest.* 124:4693–4708.
- Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22:2472–2479.
- Zhang SH, et al. 2013. Effect of muscle-fiber type on glycogenin-1 gene expression and its relationship with the glycolytic potential and pH of pork. *Genet. Mol. Res.* 12:3383–3390.
- Zou Z, Zhang J. 2015. No genome-wide protein sequence convergence for echolocation. *Mol. Biol. Evol.* 32:1237–1241.

Associate editor: Jay Storz